



Construction and Application of User Profile for Customer Management Based on RFM and K-Means

Xiaoran Li¹ and Hui Xia^{2*}

¹School of Information Engineering Jingdezhen University, Jingdezhen, Jiangxi, 333000, China, email: 2472781419@qq.com

²School of Information Engineering Jingdezhen University, Jingdezhen, Jiangxi, 333000, China
*375473215@qq.com

Abstract. To achieve effective customer management, the construction method of user portrait is studied. Taking the customers of a shopping mall as the research object, the RFM model is used to analyze customer value, and the 3 indicators of R,F,M are constructed to evaluate customer value. K-means clustering is applied to divide customers into groups, and the similarity of each group of customers is extracted to construct the user profile. Finally, 194,761 customer information data and 911,702 customer consumption data from January 2020 to January 2023 of a shopping mall are used as samples to construct user profiles. The results show that the customers of the mall can be classified into 5 categories, and the labels of their user profiles are important retained customers, ordinary retained customers, potential VIP customers, ordinary VIP customers, and important VIP customers. According to the consumption characteristics of the customers revealed by the user profile, corresponding customer management strategies are proposed to realize the purpose of tapping the potential of customers and enhancing the competitiveness of the shopping mall.

Keywords: Customer management, User profile, RFM model, Cluster analysis.

1 INTRODUCTION

With the rapid progress of society and the rapid advancement of technology, customer management has become a key element for the success of enterprises. Consumers have increasingly high expectations for service quality, and they desire more personalized and precise service experiences. The rise of e-commerce has brought new opportunities and challenges to customer management, and the intensification of competition has made retaining customers crucial. In this context, how to effectively manage customers, improve customer satisfaction and loyalty, and reduce customer churn rates has become an urgent problem to be solved [1].

Many scholars have conducted research on this issue from the perspective of combining theoretical research and practical situations. Song et al. [2] used questionnaires and interviews to study the CRM of Changchun Boze Company, explored the status of CRM, and proposed an effective CRM program. Liu et al. [3] conducted a study on a certain company in the steel trading industry. Analyzed its actual management situa-

tion in detail, constructed a customer value evaluation model with AHP, further subdivided customer groups using cluster analysis, and proposed corresponding marketing strategies. Rachel et al. [4] used the real consumption data of Tmall to classify the value of users using RFM model with K-Means clustering analysis, so as to provide personalized services for different users and realize the precision marketing based on user classification. Cai et al. [5] explored the traditional customer value evaluation methods and segmentation methods, utilized existing customers to reestablish the customer management and segmentation system, and combined the two to obtain the customer segmentation value model, and proposed the corresponding customer development strategies and product demand programs. Li et al. [6] constructed RFM model of user value using the most recent consumption, consumption frequency, and consumption amount according to the characteristics and laws of user consumption to identify the potential value of the user and construct a user portrait in order to realize precision marketing. Based on a comprehensive understanding of a recommendation system that integrates multiple attributes, Patro et al. [7] establish profiles for users, understand their potential needs, and provide accurate personalized recommendations. Liu et al. [8] pointed out that companies want to make the successful implementation of precision marketing, must be based on big data technology, the development of user profiles, accurate positioning of the customer base, master customer demand, so that the precision marketing model closer to the real needs and preferences of customers.

Based on the achievements of existing customer management, this paper collects basic customer information and consumption data of a certain shopping mall in the past three years to conduct in-depth analysis of customers; Using the RFM model to extract key value features of customers, breaking them down into different user groups based on k-means clustering, constructing a complete user profile, achieving the goal of concretizing user images, and helping shopping malls improve their quality level in refined marketing and personalized services.

2 DATA SOURCES AND RESEARCH METHODS

2.1 Data Sources

The data comes from a shopping mall and contains 194,761 pieces of customer basic data and 911,702 pieces of customer consumption data from January 2020 to January 2023. The customer basic data include: membership card number, member's birth date, gender, and member's enrollment and registration time; the customer consumption data include: bill number, product code, price, product name, sales quantity, sales amount, bonus points, time and other features, as is shown in Table 1.

Table 1. Features of customer

No	Feature	Description
1	CardID	Customer card number
2	BirthDate	Customer's birthday
3	Gender	Gender, 0 for women and 1 for men
4	RegistrationDate	Customer registration time
5	Dtime	Time of generation of customer consumption
6	ProduceID	Product code
7	SaleCount	Sales quantity
8	Price	Sales price
9	Monetary	Sales amount

2.2 Research Methods

The RFM model measure the value of customers using three key indicators: the most recent consumption time, consumption frequency, and consumption amount [9]. It is suitable for situations with a large number of customers, which is consistent with the data characteristics in this paper. The K-means method belongs to a distance based unsupervised clustering algorithm, which can reasonably divide customers into groups based on their value characteristics. The purpose of group division is to facilitate the extraction of customer similarity. User portrait refers to the virtual representative of real users, which uses the real data of users to build the model of target users [10]. User profile can depict the essential characteristics of customers and achieve refined management of customers.

3 USER PROFILE CONSTRUCTION

3.1 Data Cleaning

For the raw data, the following data cleansing is performed in Jupyter Notebook using the python pandas tool library: (1) delete the records of non-member customers and retaining the records with customer information; (2) delete the false customer records with the customer registration date and the customer's birthday outside of January 1st,1900 or January 1st, 2023 ; (3) delete the records where the customer's birthday and gender are missing; (4) delete the records where price, sales quantity, and sales amount are negative. Finally, 31,283 customers and 416,285 consumption records remain.

3.2 Calculation of RFM Indicators

Based on customer basic information data and consumption information data, calculate the R, F, and M indicators for 31283 customers. The indicator R is obtained from the time generated by customer consumption. Using January 4, 2023 as the reference date, the number of days between the last consumption time and the reference date is taken as the R value; The F-value represents the total number of times customers

spent from January 2020 to January 2023, and can be obtained by counting the consumption records of each customer; The M value is the total consumption amount of customers, which can be obtained by counting the consumption records of each customer. After calculating the R, F, and M values of all customers, basic statistical characteristics were described, and the results are shown in Table 2. Further draw the boxplots of R, F, and M, as is shown in Figure 1.

Table 2. Statistical information of R, F and M

	R	F	M
count	31283	31283	31283
mean	422.12908	12.93741	18267.03
std	375.736254	31.826072	51524.94
min	0	1	0
25%	73	2	1836
50%	312	5	5218
75%	866	12	15472.9
max	1098	3018	3257258

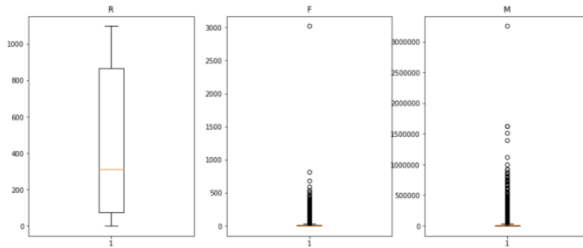


Fig. 1. Boxplots of R, F and M

3.3 K-Means Clustering

Using R, F, and M as evaluation indicators for customer value, we comprehensively utilize Python Pandas and SK-learn tool libraries to implement K-means algorithm for customer group division. Due to the fact that the effectiveness of K-means clustering is easily affected by extreme values, the identification of extreme customers should be carried out first.

Identification of Extreme Customers. Observing the boxplots of the three indicators (see Figure 1), there is a point of extreme value in the boxplots of both the F and M indicators that deviates from the other data points, and after regressing to the customer information data, it is found that the point in these two plots corresponds to the same customer, whose information is shown in Table 3. This customer has a great frequency and amount of consumption, and the recent consumption time is relatively short, and its consumption characteristics significantly deviate from the consumption

of most customers in this mall. It is possible that this customer exists, but it is not representative, so this customer is not involved in the clustering.

Table 3. Corresponding information for extreme customers

Feature	CardID	BirthDate	Gender	RegistrationDate	age	R	F	M
value	7ff6682d	22757	0	42998	57	2	3018	3257260

Determination of the Number of Clusters. Use elbow method and silhouette coefficient method to select the number of clusters (K), and comprehensively consider the results of both to determine the optimal K. The elbow method characterizes the variation of the sum of squared errors within clusters (SSE) as K increases. The silhouette coefficient method characterizes the variation of silhouette coefficients as K increases. SSE and silhouette coefficient are both indicators for evaluating clustering effectiveness. The changes in both are shown in Figure 2. When $K > 5$, the decrease in SSE and silhouette coefficient tends to be gradual, no significantly decrease. Therefore, $K=5$ is chosen as the optimal number of clusters, and dividing the customer group into 5 categories is the best.

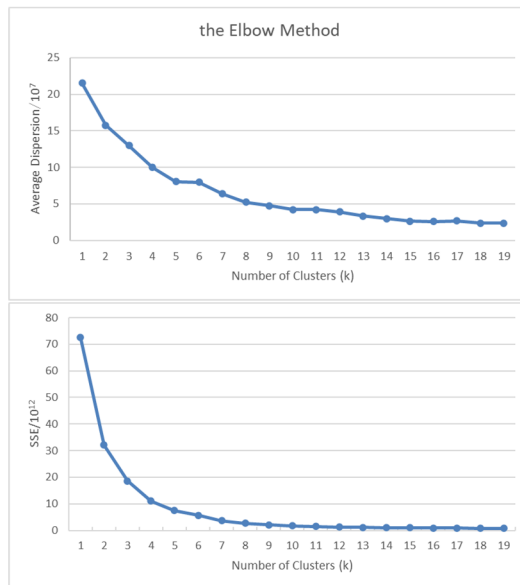


Fig. 2. Elbow diagram and silhouette coefficient diagram

Results of Clustering. The optimal number of clusters is determined to be 5, and K-means clustering is performed on 31,282 customers to obtain the clustering results. It can be known that the clustering centroids of 5 groups of customers are (8.81,318.28,853444), (34.19,172.16,346235.66), (56.52,86.65,153451.9), (167.8,34.95,51788.84), and (468.72,6.35,6562.43), and according to the cluster-

ing labels for each group of customers' visits and metrics. As in Table 4 (sorted by the number of people in the class in descending order).

Table 4. Final clustering centers

Feature	R	F	M	Customers counts	percentage
Class 1	8.8125	318.2813	853444.0020	32	0.10%
Class 2	34.1869	172.1566	346235.6560	198	0.63%
Class 3	56.5184	86.6483	153451.9020	760	2.43%
Class 4	167.8002	34.9492	51788.8374	3452	11.04%
Class 5	468.7163	6.3509	6562.4276	26840	85.80%

When cluster analysis is performed and the number of customers is set to 5, the customers can be successfully categorized into 5 classes and the number of customers in the 5 classes is shown in Figure 3. The number of customers in class 1, class 2, class 3, class 4 and class 5 are 32, 198, 760, 3452 and 26840 respectively.

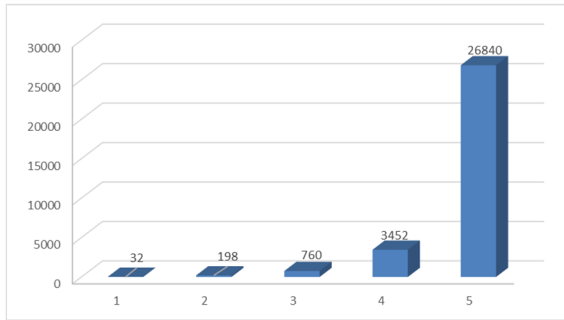


Fig. 3. Number of customers in each group

3.4 User Profiling

According to the above clustering results, customers can be divided into 5 groups, and the clustering center of each group is the average level of R, F, and M. The range of R, F, and M for each group of customers can be obtained from the minimum and maximum values of the three indicators of R, F, and M for each group of customers. The results are shown in Table 5.

Furthermore, it was found that gender and age have a certain impact on customer value through statistical analysis of basic customer information and consumption data. But the distribution of gender and age among the 5 groups of customers is highly similar, that is, gender and age are not the influencing factors that lead to differences between customer groups. Therefore, gender and age factors are not considered characteristics in user profiles.

Table 5. Range of R, F and M among 5 classes of customers

class	R	F	M	Number of males	Number of females
class 1	0-55	6-818	605287.4-1620435	2	30
class 2	0-913	2-594	250520.55-587580.12	24	174
class 3	0-988	2-412	102758.4-248002.64	95	665
class 4	0-1097	1-330	29186.1-102586.5	508	2944
class 5	0-1098	1-104	0-29172.1	3723	23117

Draw a user profile based on the average level and range of R, F, and M indicators for each group of customers. In this process, it is believed that the R, F, and M indicators are equally important to avoid bias caused by excessive emphasis on a specific indicator. The profiles of the five types of customers are described as follows.

For type 1 customers, the profile can be described as: the last consumption time was very short; High frequency of consumption; The consumption amount is very large. It is known that this type of customer has high loyalty, satisfaction, strong purchasing power, and extremely high value, and is defined as an important VIP customer.

For type 2 customers, the profile can be described as: the most recent consumption time is relatively short, the consumption frequency is relatively high, and the consumption amount is relatively high. It is known that this type of customer has high loyalty, satisfaction, and purchasing power, and has high value. Its label is defined as a regular VIP customer.

For type 3 customers, the profile can be described as: the most recent consumption time is relatively short, the consumption amount is relatively high, but the consumption frequency is not high. Infer that this type of customer has low loyalty but great potential, and define it as a potential VIP customer or important development customer.

For type 4 customers, the profile can be described as: the most recent consumption was relatively long, and the frequency and amount of consumption were average. This type of customer has low loyalty, low satisfaction, and average purchasing power, and is defined as a key retained customer.

For type 5 customers, the profile can be described as: the most recent consumption was far away, the frequency of consumption was low, and the amount of consumption was not high. This indicates that this type of customer has low loyalty and is likely to have been lost, with minimal spending power. Its label is defined as a regular retained customer.

4 APPLICATION OF USER PROFILES

The overall design principle for customer management based on user profiles is to stabilize the existing customer base, expand the number of VIP customers, promote the gradual transformation of the other three types of customers to VIP customers, and achieve the upgrading of customer level[11]. Based on the value characteristics of

user profiles, optimize customer relationship differentiation management, and propose targeted solutions as follows:

(1) VIP customers are the core customers of the mall and also the guarantee for its future development. Choose to strengthen the value management of such customers from both marketing and service aspects.

First, Build exclusive marketing strategies, grant special preferential policies and discount privileges, design personalized marketing plans and customized promotional activities; Organize high-end membership activities to enhance their sense of respect and belonging; Create a rapid response mechanism to ensure that VIP customer orders can be prioritized and efficiently executed; Regularly conduct sales follow-up work, understand their needs and feedback, and adjust sales strategies in a timely manner.

Second, Configure a professional sales team to provide one-on-one attentive service to important customers; Establish a dedicated service channel for VIP customers, providing 24-hour online service to ensure they can solve problems at any time; Provide unexpected service experiences, such as value-added services, priority after-sales services, etc; Regularly care and greet VIP customers, and offer customized gifts and blessings on special holidays and anniversaries.

(2) For potential VIP customers, appropriate strategies should be adopted to increase loyalty and enable them to develop into VIP customers.

First, In terms of marketing, utilizing online channels for precise marketing push, delivering valuable information and personalized content to customers; Plan brand stories and cultural dissemination to generate deeper identification and resonance among customers towards the brand; Establish long-term purchasing incentive mechanisms such as cumulative consumption rewards; Carry out corresponding promotional activities to promote customer consumption.

Second, in terms of service, provide proactive services, such as regular follow-up visits. Set up a customer service team to ensure efficient communication and problem-solving, and deepen emotional connections with customers. Conduct customer satisfaction surveys and continuously improve service quality based on feedback. Continuously monitor changes in customer needs and adjust service strategies in a timely manner.

(3) In the face of retaining customers, especially important ones, it is necessary to deeply analyze the corresponding user profile, explore the basic reasons for customer loss, analyze the shortcomings in customer marketing and service, and make improvements, such as conducting customer retention data analysis, arranging specialized sales personnel to conduct follow-up visits on customer retention, and identifying service loopholes for lost customers.

Implement customer management strategies based on user profiles. In the short term, by comparing the average changes in R, F, and M for each type of customer, the effectiveness of management strategies can be judged, facilitating timely and dynamic adjustment of management strategies. Besides, user profiles may change over time, and shopping malls should regularly update customer profiles, such as redefining user profiles on an annual basis. By changing the customer profile, timely adjusting marketing and service measures for customers, achieving the goal of optimizing customer management and improving the competitiveness of the mall.

5 CONCLUSION

To develop a refined management plan for the shopping mall, customer consumption details of the last 3 years are selected as sample data, and RFM model and k-means models are used to obtain user profiles.

(1) Based on the characteristics of the user profile, divide the mall's customers into 5 categories. Through quantitative indicators to evaluate and rank the value of customers, scientific classification of customers can be achieved, laying a solid foundation for optimizing customer management in shopping malls.

(2) Based on the above user profiles, Carry out customer management and propose appropriate management strategies for different types of customers with the help of the characteristics of each type of customer in the user profiles.

(3) Three classic indicators are selected for the evaluation of customer value in this paper, more indicators can be further considered to make the result more reasonable and scientific; There are multiple ways to define the scope of the three indicators R,F and M in user profiles, resulting in different outcomes, that is the problem needs to be carefully considered in future.

REFERENCES

1. Zhang, H.: The Application of Customer Management in Enterprise Marketing [J]. *Modern Marketing (Operation Version)*, (04): 42-43 (2021). DOI: 10.19921/j.cnki.1009-2994.2021-04-0042-021.
2. Song, L.: Research on Customer Relationship Management of Changchun Brose Automotive Components Co., Ltd. [D]. Jilin University (2023). DOI: 10.27162/d.cnki.gjlin.2023.004762.
3. Liu, N.: Research on Customer Value Analysis of Company A Based on Cluster Analysis [D]. Shandong University (2021). DOI: 10.27272/d.cnki.gshdu.2021.001738.
4. Liu, R., Song, Z.: Analysis of Tmall User Value Based on RFM Model [J]. *Economist*, (12): 243-244 (2022).
5. Cai, Q.: Value-Based Customer Segmentation and Promotion Strategy of BLN Company [D]. University of Electronic Science and Technology of China (2019).
6. Li, J., Zhao, G.: Research on the Construction of E-commerce User Profile Based on Big Data [J]. *E-commerce*, (01): 46-49 (2019). DOI: 10.14011/j.cnki.dzsw.2019.01.020.
7. Patro, S. G. K. , Mishra, B. K. , Panda, S. K. , et al.: Hybrid Social Recommender Systems for Electronic Commerce: A Review[C]// 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (2020).
8. Liu, H., Lu, H., Ruan, J., et al.: Research on the precision marketing segmentation model based on the mining of "user portrait" [J]. *Journal of Silk*, 52(12): 37-42 + 47 (2015).
9. Soeini, A. R., Fathalizade, E.: Customer Segmentation based on Modified RFM Model in the Insurance Industry.Proceedings of The 2012 4th International Conference on Machine Learning and Computing.Ed.Industrial Development Renovation organization of Iran,Director Development Renovation Investment Tehran,Iran;Department of e-commerce, Nooretouba University Master's degree student,Rasht,Iran;104-107 (2012).
10. Cooper: *The Inmates Are Running the Asylum* [M]. Beijing: Publishing House of Electronics Industry, 10 (2006).

11. Specchia, A.: Customer Relationship Management (CRM) for Medium and Small Enterprises: How to Find the Right Solution for Effectively Connecting with Your Customers[M]. Taylor and Francis (2022). DOI:10.4324/9781003148388.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

